CENTER FOR DRUG EVALUATION AND RESEARCH

Application Number 21-289

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

NDA#: 21-289

Applicant: Ferring Pharmaceuticals, Inc.

Name of Drug: Ovanex (purified FSH)

Documents Reviewed: Vol 10B, Vol 10D dated September 28, 2000

Medical Officer: Ridgely Bennett, M.D., HFD-580

Background

The sponsor has submitted two **unblinded**, active-controlled clinical trials, one in support of Ovanex's efficacy for *in-vitro* fertilization (99-04) and the other for Ovulation Induction (99-03) in anovulatory or oligo-ovulatory infertile females. Each trial compared purified FSH SC and purified FSH IM to Follistim SC (active control).

Trial 99-03 (Ovulation Induction)

Patients were treated for one cycle. The primary efficacy endpoint was the percentage of patients who ovulated. **hCG was given at the investigator's choice rather than patients being randomized to hCG or not**. According to the sponsor, the tolerable non-inferiority margin relative to Follistim was an absolute 25% difference in the percentage of ovulating women. The absolute difference is derived from an earlier version of the protocol which specified a *relative* 35% margin. Since the anticipated ovulation incidence was 70% in the Follistim group, this corresponded to an absolute difference of 25% (.35 X 70%=25%).

Sponsor's Results

A total of 111 women were randomized among 14 clinical sites: FSH SC: N=36, FSH IM: N=37 and Follistim: N=38. The percentages of women who ovulated were that 25/36 (69.4%) for FSH SC, 26/37 (70.3%) for FSH IM, and 30/38 (78.9%) for Follistim. The sponsor's confidence intervals for the differences in percentages between the two test groups FSH SC, FSH IM and the Follistim control group were (-26.2%, 7.2%) and (-25.1%, 7.8%), respectively. This reviewer used a large sample binomial software program which produced lower bounds of -29% and -28%, respectively.

However, when only the women who received hCG are considered, a very different picture emerges. The percentages of ovulation were 25/26 (96.2%), 26/28 (92.9%), and 30/35 (85.7%) in the FSH SC, FSH IM and Follistim groups, respectively. The respective lower bounds of the confidence intervals for the difference between the test groups and the Follistim group were -1.1% and -5.5%, respectively.

Discussion

The results of the ITT analysis suggest that FSH may be inferior to Follistim, but there is a plausible explanation due to bias in favor of Follistim. As an illustration, take the comparison involving the FSH SC group. First, of the 13 women on FSH SC or Follistim) who did **not** receive hCG, **none** ovulated. Of those 13, 10 were in the FSH SC group. This means that, from the data in the trial, one would estimate the probability of ovulating without hCG to be zero. Thus, one cannot control for the clear imbalance in hCG exposure even if one wants to regard the data as observational and not randomized.

In order to track the source of the imbalance, this reviewer looked at the groups' exposure to hCG by clinical site and then, for those which showed a substantial imbalance, inspected the ovulation results between the two groups. There were two cases in which the ovulation results by group were *exactly the same* as the hCG exposure (yes or no) by group. Specifically, at site #1, 2 of the 4 FSH SC women ovulated while 4 of the 5 Follistim patients did. Those fractions are the same as those who received Follistim in each group. At site #14, 1 out of 3 FSH SC women ovulated, while 3 out of 3 Follistim patients did. Again the fractions were the same with respect to hCG exposure.

An alternative approach is to assume that receiving hGC is independent of treatment group. If that is the case, then one could estimate the *additional ovulations that would have occurred had the women received hCG.* Using the pooled estimate of the probability of ovulating given that a woman received hCG (.90) and applying that to the 10 women who did not receive hCG in the FSH SC group and the corresponding 3 women in the Follistim group, the projected fraction of women who would have ovulated in the ITT FSH SC group is 34/36, while that in the Follistim group is 33/38. The lower bound of the (albeit artificial) "confidence interval" is -5.1%, clearly within the margin of non-inferiority. This approach would be invalidated if, in fact, reasons for withholding hCG is related to the action of the particular drug being administered. For example, if a clinician thought or knew that women who was taking FSH SC was somehow less likely to ovulate even if they received hCG, then the information about the imbalance in receiving hCG would be informative. **Recall that the study was not blinded.** However, in the absence of such a hypothetical interaction between drug and exposure to hCG (whether pharmaceutically based or due to investigator bias), a reasonable sensitivity analysis is to project the number of ovulating women from those who did not receive hCG.

The same pattern occurred comparing FSH IM to Follistim in that of the 12 women who did not receive hCG, none ovulated and 9 were in the FSH IM group.

Conclusion

The fact that women were not randomized to hCG and the lack of ovulation in women who did not receive hCG make any formal analysis which "corrects" for any imbalance in hCG exposure impossible in this trial. Either one has to restrict inference to those women who received hCG or one can project the number of women who would have ovulated had they received hCG. In the former case, results appear to support the non-inferiority of FSH SC and FSH IM relative to Follistim for induction of ovulation in conjunction with administration of hCG.

Trial 99-04 (IVF)

A total of 177 women were randomized among 11 clinical sites to FSH SC (N=60), FSH IM (N=59), and Follistim N=58). The **ITT group** patients were those who received medication. Patients not receiving hCG were assigned "zero' for efficacy endpoints. **The primary efficacy outcome was the number of oocytes produced. The primary efficacy responder subgroup** (N=167) consisted of patients who had received hCG and underwent oocyte retrieval: N=56:SC, N=55:IM, and N=56: Follistim. In the respective treatment groups, there were 4, 4, and 2 women who did not receive hCG.

The study was originally planned to enroll 44 patients per group in order to have an 80% chance that a two-sided 95% confidence interval would "detect a change in the number of oocytes of 1.2.. The primary analysis was ANCOVA with age and BMI index as covariates. The protocol does not state this 1.2 as an explicit non-inferiority margin, but rather as a consequence of a power calculation using 44 patients per group. The minutes of a meeting with the company on April 24, 2000 states that the sponsor should state the statistical hypotheses to be tested. As just mentioned, such a statement would have to include a statement of a clinical margin to be excluded, not just a statement of the properties of a test given the planned sample size. In addition, the minutes state that "the Division takes the worst case scenario to be that Follistim produces a mean of at least 1.2 more oocytes than either delivery method of FSH."

The original protocol was modified in a subsequent submission (September 14, 2000) which restated the original properties concerning the 1.2 oocyte difference from Follistim, but also included a revised calculation which states that "there is an 80% power to *detect a relative difference of less than 30%* if the number of oocytes retrieved in the reference group is 10 with 50 evaluable patients in each group" (emphasis added). Again, there is no statement of a clinical margin. The 30% figure was included in a modification dated March 29, 2000 and then again on June 9, 2000. However the date of the submission is September 14, 2000, just days before the submission of the NDA. As mentioned below, it was only *afterwards in the study report*, that there was any mention of a 20% clinical margin as an non-inferiority standard, seemingly having nothing to do with the 30% stated in the protocol modifications. This a wholly unsatisfactory situation. The submission of the revised protocols only shortly before submission of the NDA raises the possibility that the revisions were driven by examination of unblinded data. The 20% margin mentioned in the study report simply means that the sponsor has failed in its job to provide adequate *a priori* means to interpret the results of the trial.

Sponsor's Results

The main results are that in the **ITT group** (for which the 10 non-hCG patients were assigned "zero oocytes"), the mean number of oocytes per woman were 13.3, 12.2, and 13.1 in the FSH SC, FSH IM, and the Follistim groups, respectively. The respective standard errors were 1.05, 1.06, and 1.07. The two-sided 95% confidence intervals for the mean differences FSH SC-Follistim and FSH IM - Follistim using Dunnett's critical values in order to preserve overall Type I error at 5% were (-2.9 oocytes, 3.4 oocytes) and (-4.1 oocytes, 2.3 oocytes), respectively. For *mature* oocytes, the confidence intervals were (-1.9, 2.6) and (-3.0, 1.5), respectively.

In the **primary responder subgroup**, the respective numbers were 14.3, 13.1, and 13.6 in the FSH SC, FSH IM, and the Follistim groups, respectively with standard errors of 1.03, 1.04, 1.03. The 95% confidence intervals for the mean differences FSH SC- Follistim and FSH IM - Follistim using Dunnett's critical values in order to preserve overall Type I error at 5% were (-2.4 oocytes, 3.8 oocytes) and (-3.6 oocytes, 2.6 oocyes), respectively. For *mature* oocytes, the confidence intervals were (-1.5, 2.8) and (-2.7, 1.7), respectively.

Discussion

The report states that one of the goals of the trial was to show that the mean number of oocytes on FSH is within 20% (relative) of Follistim's mean. **This 20% was never stated in any protocol.** In order to "test" this hypothesis, the sponsor essentially divided the lower bound of a confidence interval for the difference in mean oocytes by the *observed* mean number of oocytes in the Follistim group. However, the sponsor evidently used the lower bound of a *one-sided 95% confidence interval* in one part of the submission but also provided two-sided intervals in another part. This review uses only the two-sided 95% confidence intervals. For instance, the Follistim group has a mean of 13.1 oocytes and the lower bound for the ITT FSH SC group was -2.9 (two-sided). Thus, since 2.9/13.1=.22 which is greater than 20%, the sponsor would be bound to say that their test failed the non-inferiority standard.

This method of using the *observed data* as a basis to apply a 20% standard is not correct. The actual question being asked (recall that the sponsor never stated a well-defined hypothesis concerning a comparison of means of the *population of potential users*, which is the point of hypothesis testing to begin with) concerns the ratio of the *population means* of the test arms (FSH SC or IM) to the mean of the standard (Follistim). Thus if μ_T is the mean of the test arm and μ_F is the mean of the Follistim (standard) arm, then what we really want to know is whether the lower bound of a 95% confidence interval for this ratio of population means (μ_T/μ_F) is less than 80% which is the same as saying that the relative difference could be greater 20%.

As a *post hoc* analysis, this reviewer used Fieller's theorem to calculate such confidence intervals, this reviewer found that the lower bound of a two-sided 95% confidence interval for the ratio of the FSH SC mean to the Follistim mean to be 82% without Dunnett's correction and 80% with Dunnett's correction. For FSH IM, the lower bound of the Dunnett's confidence interval is 72%.

An interesting feature of the plan is that the sponsor expected the standard deviation in each group to be 2 oocytes when, in fact, the trial data estimates the standard deviation to be more like 8 oocytes.

The sponsor's analysis does not stratify on center in the analysis. This reviewer has checked results with 'site' in the model and the results are similar to those of the sponsor.

Conclusion

The sponsor's protocol submissions have not provided an unambiguous way to assign a non-inferiority margin for this trial. However, the 95% two-sided confidence intervals for the mean differences provided by the sponsor are statistically valid as are those that this reviewer produced using Fieller's theorem.

/S/

David Hoberman, Ph.D. 7/25/01
Mathematical Statistician

Concur: Dr. Kammerman

Dr. Nevius

cc:
NDA# 21-289
HFD-580
HFD-580/DSpell-LeSane, RBennett, SSlaughter
HFD-715/CAnello, DHoberman, LKammerman

APPEARS THIS WAY
ON ORIGINAL